

Abstract

Why data should be available

Science works more surely and more rapidly with an open and transparent process. Part of that process is the ability to criticize. Criticism is more effective with access to data.

Analysis of medical observational studies has well-recognized problems: multiple testing, multiple modeling, and bias. Less well-recognized is the real world impact of researcher incentives to obtain “statistically significant” results. Published papers are necessary for grants and fame. There is a need to find analysis and managerial strategies to counter “researcher initiative”.

Access to study protocol, statistical analysis code and electronic data sets to be able to replicate published claims is a good step toward valid claims.

Data Availability



S. Stanley Young

Assistant Director for Bioinformatics
National Institute of Statistical Sciences

1. I consume data sets.
2. I invent statistical methods.
3. People trust me to do an honest analysis.
4. People expect valid claims.

Funding Agencies

1. Making data available is a requirement.
2. Separated data generation from analysis.
3. Require “Reproducible Research”.
4. Repository: protocol, analysis code, data.
5. Fund analysis, re-analysis, replicate studies

The Worker is not the Problem

W. Edwards Deming,

the most visionary innovator ever on quality control, said

The worker is not the problem.

The problem is at the top! Management!

To Deming, blaming the workers—individual researchers—is as incorrect as it is useless.

Bringing the system under control is the responsibility of those managing it, [editors and funding agencies](#).

Air Pollution and IQ

PEDIATRICS[®]

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

ImpFac = 5.4

Prenatal Airborne Polycyclic Aromatic Hydrocarbon Exposure and Child IQ at Age 5 Years

Frederica P. Perera, Zhigang Li, Robin Whyatt, Lori Hoepner, Shuang Wang, David Camann and Virginia Rauh

CONCLUSION: These results provide evidence that environmental PAHs at levels encountered in New York City air can affect children's IQ adversely. *Pediatrics* 2009;124:e195–e202

8 outcome variables, 8 chemicals, 7 covariates

Search Space: $8 \times 8 \times 2^7 = 8,192$

NIEHS funded; refused to provide data set.

Observational study claims tested in Randomized Clinical Trials

ID#	Pos	<u>Neg</u>	#Claims	Treatment(s)
				Claims based on observational study
1	0	1	3	<u>Vit E</u> , beta-carotene
3	0	3	4	Hormone Replacement <u>Ther.</u>
5	0	1	2	<u>Vit E</u> , beta-carotene
6	0	0	3	<u>Vit E</u>
10	0	0	3	Low Fat
11	0	0	3	<u>Vit D</u> , Calcium
12	0	0	2	Folic acid, <u>Vit B6</u> , B12
13	0	0	2	Low Fat
14	0	0	12	Vit C, <u>Vit E</u> , beta-carotene
17	0	0	12	<u>Vit C</u> , <u>Vit E</u>
18	0	0	3	<u>Vit E</u> , Selenium
new	0	0	3	<u>HRT+antioxidant vits**</u>
	0	5	52	

Why claims fail: Big Three, +

1. Bias

2. Multiple testing

3. Multiple modeling

A worker goal is $p < 0.05$.

Feynman, 1997

In summary, the idea is to try to give all the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.

Feinstein, Science 1988

All of these highly publicized accusations of "menace" came from research that had been approved by the "peer review" of authoritative experts. The peer review process, however, provides assurance only that an act of research complies with accepted methods in a field of investigation.

Scientific Standards in Epidemiologic Studies of the Menace of Daily Life

ALVAN R. FEINSTEIN

Separate data collection and cleaning from analysis.

1. Data collection and cleaning can take over 80% of researcher efforts.
2. Data staging can bias results.
3. Stop sneak analysis tuning.

Funding Agencies

1. Making data available is a requirement.
2. Separated data generation from analysis.
3. Require “Reproducible Research”.
4. Repository: protocol, analysis code, data.
5. Fund analysis, re-analysis, replicate studies

Contact Information

Stan Young

young@niss.org

<http://www.niss.org/content/s-stanley-young>

919 685 9328